



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil dari implementasi dan hasil uji coba sistem, simpulan dari penelitian adalah sebagai berikut.

1. Implementasi metode Extreme Gradient Boosting (XGBoost) pada *sentiment analysis* dalam mengklasifikasikan komentar menggunakan *word embedding* telah selesai dibuat. *FastText pre-trained model* yang digunakan merupakan kumpulan kata bahasa Indonesia dan didapatkan dari laman resmi *FastText* dan model lainnya dibuat menggunakan *dataset* Pemilu2014 yang digabungkan dengan *dataset* SentimentTwitter.
2. Hasil dari uji coba dengan menjadikan *F1-score* sebagai acuan dalam pemilihan model terbaik, didapatkan model terbaik ketika menggunakan *dataset* gabungan Pemilu2014 dan SentimentTwitter dan *FastText pre-trained* model sebagai *embedding* model dengan skenario pemberian bobot pada setiap *class* dimana *class* negatif menjadi acuan (*negative class weighting*). *Dataset* SentimentTwitter ditambahkan sebagai langkah perbaikan dalam menangani ketidakseimbangan jumlah data dalam *dataset* Pemilu2014. Hasil yang diperoleh yaitu akurasi sebesar 74.8%, *precision* sebesar 50%, *recall* sebesar 48%, dan *F1-score* sebesar 49%.

5.2 Saran

Berdasarkan penelitian yang dilakukan, terdapat beberapa saran untuk pengembangan lanjutan, yaitu sebagai berikut.

1. Pengembangan selanjutnya diharapkan dapat menyempurnakan proses klasifikasi dengan melakukan *pre-processing* yang lebih bervariasi. Penggunaan *hyperparameter* XGBoost lainnya. Diharapkan dengan itu, performa yang dihasilkan bisa meningkat dan lebih memuaskan.
2. Menggunakan *dataset* dengan tingkat keseimbangan yang relatif tinggi. Diharapkan dapat memberikan hasil yang lebih baik dan hasil performa yang lebih memuaskan dalam melakukan klasifikasi *multiclass*.
3. Mencoba menggunakan metode *word embedding* lain seperti *GloVe* dan *Word2Vec* sebagai *feature extraction*. Dengan mencoba data *pre-trained* lainnya dan dengan bantuan *library* dari *gensim* ataupun bawaan dari *word embedding* tersebut diharapkan memberikan hasil performa yang lebih baik dan memuaskan.
4. Mencoba memperbesar *range value* dan memberikan variasi *hyperparameter* lainnya di *RandomizedSearchCV* dan mencoba metode *cross validation* yang lain. Diharapkan dengan memperbesar *range value* dan memberikan variasi *hyperparameter* dapat memberikan hasil performa terbaik.